International Journal of Advances in Data and Information Systems
Vol. 4, No. 2, October 2023, pp. 125~134
ISSN: 2721-3056, DOI: 10.25008/ijadis.v4i2.1297

Hybrid Model Transfer Learning ResNet50 and Support Vector Machine for Face Mask Detection

Eko Agus Moh. Iqbal[1], Ririen Kusumawati[2], Irwan Budi Santoso[3]
[1,2,3] Faculty of Science and Tecnology, Program Specification for Master Study in Computer, Universitas Islam Negeri Maulana Malik Ibrahim, Indonesia

Article Info

Article history:

Received Jun 05, 2023
Revised Jul 16, 2023
Accepted Sep 03, 2023

Keywords:

Mask Detection
HSV
Transfer Learning
ResNet50
Support Vector Machine

ABSTRACT

The Covid-19 virus caused a health crisis in Indonesia. This virus is so deadly that it has caused many fatalities which have caused the whole world including the government to pay major attention to the Covid-19 pandemic. The Indonesian government has issued several policies to prevent the spread of this epidemic, one of which is wearing a mask in public places. One approach that is widely used in the field of computer vision is the Convolutional Neural Network (CNN) transfer learning. In this study, Hybrid Model Transfer Learning ResNet50 and SVM with RGB to HSV preprocessing is presented to detect masks in facial images. This model consists of three process components. The first is preprocessing RGB images to HSV, the second component is for Feature Extraction with ResNet50 and the third is mask classification on face images with Support Vector Machine (SVM). From dataset of 7328 training and testing data were carried out. The first model, without preprocessing the image data with ResNet50, produces an accuracy of 86.52%. The second model, the model with preprocessing converts image data from RGB to HSV with ResNet50 resulting in an accuracy of 99.18%. In the third model, without preprocessing with ResNet50 and SVM which has an accuracy of 90.55%. The fourth model, the model with preprocessing converts image data from RGB to HSV with ResNet50 and SVM resulting in an accuracy of 98.36%.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Eko Agus Moh. Iqbal,
Faculty of Science and Tecnology,
Program Specification for Master Study in Computer Science (Magister Informatika)
Universitas Islam Negeri Maulana Malik Ibrahim, Indonesia
Jl. Gajayana No.50, Dinoyo, Kec. Lowokwaru, Kota Malang, Jawa Timur 65144.
Email: eamiqbal19@gmail.com

1. INTRODUCTION

Digital transformation in the health sector has been carried out both clinically and non-clinically. Technology can be used to reduce direct contact because viruses can spread quickly through direct contact. Disease outbreaks that infect the respiratory tract caused by viruses are called Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV-2) [1].This virus is so deadly that it has caused many fatalities which have caused the whole world including the government to pay major attention to the Covid-19 pandemic.

Indonesia as one of the countries affected by the spread of this virus has taken several countermeasures against the spread of the SARS-Cov-2 virus. Areas that are included in the red zone of the spread of the virus [2]. In addition, he issued a circular letter Number

HK.02.02/I/385/2020 of 2020 concerning the use of masks and the provision of Handwashing Facilities with Soap (CTPS) to prevent transmission of Covid-19. One of the oldest things in the protocol is requiring every individual who has activities outside the home to wear a mask. In some cases, a person is not aware that they are infected and can pass the disease on to others. With a mask detection system, it can detect people who do not comply with the use of masks, making it possible to find potential cases early and take the necessary steps to prevent further spread.

One approach that is widely used in the field of computer vision today is the Convolutional Neural Network (CNN) as an algorithm in Deep Learning which has great popularity in terms of image classification. The CNN algorithm has several architectures such as Alexnet, ResNet, VGG, EfficientNet, and so on. In the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition in 2015, ResNet is the CNN architecture with the lowest error rates compared to other CNN architectures. This architecture has a good performance in the ILSVRC competition [3]. ILSVRC is an annual competition for image classification with various CNN architectures. The SVM method is a classification method because it is able to classify with high-dimensional data and provides experimental results that show effectiveness. In addition, SVM is capable of performing separate non-linear data pattern recognition [4].

Face mask detection using CNN, including VGG19, Xception, MobileNetV2 which were then classified using KNN and SVM. In this study, experiments were carried out between feature extraction methods and paired classification methods such as VGG19 with KNN accuracy reaching 96.65%, then MobileNetV2 with KNN accuracy of 94.92%, then Xception with SVM accuracy reaching 94.57% and MobileNetV2 with SVM produced the highest accuracy among the other pairs, reaching 97.11% [5]. MobileNetV2 and SSD architecture to detect the number of people for long distances and detect masks on faces in real time. The resulting accuracy reaches 91.7% [6].

Trained MobileNet generates feature maps. Then the Global Polling Block multi-dimensional feature map is converted into a one-dimensional vector that has 64 features. The accuracy of this method produces 99% of the actual data, and 100% for the simulation data [7]. Proposed a hybrid model, namely preprocessing using Resnet50 which is an image feature extraction model, and after that using the Decision Tree, SVM, and Ensemble algorithms for the face mask classification process. The author presents a comparison between the Decision Tree, SVM, and Ensemble Algorithms using three data sets where the classification using SVM is superior to the Decision Tree and Ensemble Algorithms. The accuracy of each of these datasets is Real-world Masked Face Dataset (RMFD) reaching 99.64%, Simulated Masked Face Dataset (SMDF) reaching 99.49% and Labeled Faces in the Wild (LFW) reaching 100% [8].

The detection of masks on faces in this study was developed using the MobileNetV2 algorithm which is used for image classification. With MobileNetV2, the built model will be used for the prediction/classification process. Of the models tested in 25 cities in Indonesia. for the highest accuracy reaching 82.76% for the city of Jambi and the lowest reaching 64.14% for the city of Surabaya [9]. Proposed using the CNN method for face mask detection. Each augmented input image passes through a convolutional layer and is defined using softmax, assigning values between 0 and 1. The results of this research achieved an accuracy of 96.90% in classifying whether a face is wearing a mask or not wearing a mask [10].

Proposed building a real-time face mask detection model using CNN in their research. The detection is performed automatically in public spaces such as schools, markets, offices, etc. The authors utilized the Keras algorithm to detect whether a person is wearing a mask or not. For classification, they employed the MobileNet architecture and OpenCV algorithm. This model serves as the backbone and is trained using TensorFlow. The accuracy achieved in this study is 97.75% [11].

Conducted research on face mask detection in public areas to reduce the spread of Covid-19. The proposed technique efficiently handles occlusion in crowded situations by utilizing one and two-stage detectors in pre-processing. The models used in this pre-processing stage are ResNet50, AlexNet, and MobileNet. ResNet50 is used to detect objects such as the neck and head, while AlexNet and MobileNet are used to achieve the best results for mask detection on faces. The proposed technique achieves an accuracy of 98.2% [12].

In those studies, preprocessing was not performed on the input images, although converting the image from RGB to HSV can handle more complex scenarios much better, which can have a significant impact on the performance of the methods used [13]. Additionally, adding preprocessing to the image, such as changing the colorspace from RGB to HSV, can save training time and improve the accuracy of the VGG19 transfer learning model [14]. Therefore, in this study, the authors used a Hybrid model of Transfer Learning ResNet50 with Support Vector Machine (SVM) for mask detection on facial images. The input data undergoes preprocessing by converting the colorspace from RGB to HSV, and then the Transfer Learning ResNet50 model is used for feature extraction, with SVM for the classification process. The use of ResNet50 in feature extraction is effective in capturing and representing complex patterns and features in images. Similarly, SVM is capable of high-dimensional data classification and is less prone to overfitting, making its classification more efficient and effective. Therefore, Hybrid ResNet50 with SVM can provide higher accuracy values with lower errors [5], [8], [12].

2. RESEARCH METHOD

The research phase begins with data collection, where the data used can be categorized as secondary data. The data is obtained from previous research data available at https://github.com/prajnasb [15], consisting of 1376 images. This data is then supplemented with additional data of 128x128 pixels dimension from https://kaggle.com [16]. Therefore, the total number of data for the research is 7328. After data collection, the system is designed, followed by experiments using several model combinations, including RGB to HSV - ResNet50 and RGB to HSV - ResNet50 - SVM. Subsequently, the results are evaluated and analyzed. The procedure of the research process is illustrated in Figure 1.

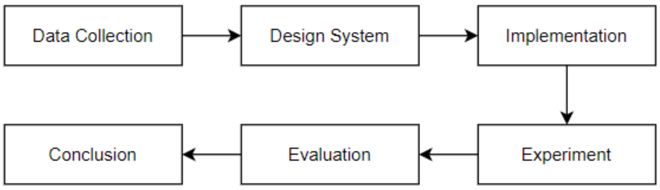

Figure 1. Research Design

The design process of the introduced model system comprises three main components. The first component is preprocessing, which involves converting the RGB image to HSV. The second component is the Transfer Learning model ResNet50, which acts as the feature extractor. The third component is machine learning, specifically SVM. In the classification process, the last layer of ResNet-50 is removed and replaced with SVM for classification, aiming to enhance the model's performance in this research. The system design is depicted in Figure 2.

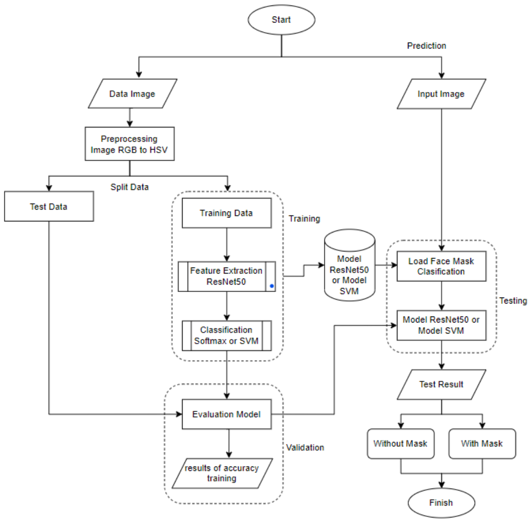

Figure 2. Design System

In Figure 2, there is a preprocessing process for the facial image dataset, namely RGB to HSV. This is part of the author's efforts to improve the accuracy of the ResNet50 model. Adding preprocessing to the image, such as segmenting the image into HSV, can save training time and improve the accuracy of the VGG19 transfer learning model. The process of converting from RGB to HSV has done using the following formula:

$$r = \frac{R}{(R+G+B)},\ g = \frac{G}{(R+G+B)},\ b = \frac{B}{(R+G+B)} \tag{1}$$

$$V = \max(r, g, b) \tag{2}$$

$$S = \begin{cases} 0, & V = 0 \\ 1 - \frac{\min(r,g,b)}{V}, & V > 0 \end{cases} \tag{3}$$

$$H = \begin{cases} 0, & if\ S = 0 \\ \frac{60*(g-b)}{S*V}, & if\ V = r \\ 60 * \left[2 + \frac{b-r}{S*V}\right], & if\ V = g \\ 60 * \left[4 + \frac{r-g}{S*V}\right], & if\ V = b \end{cases} \tag{4}$$

As an example of changing an RGB image to HSV, it can be seen in Figure 3.

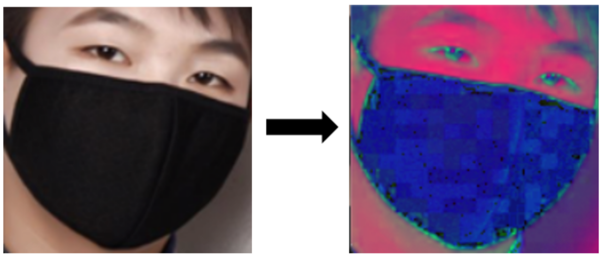

Figure 3. Convert RGB to HSV

After these image data are converted from RGB to HSV, it is then transformed into an array and stored according to their respective classes, whether wearing a mask or not wearing a mask. Next, the data will be processed using the ResNet50 model and classified using the Softmax Activation function.

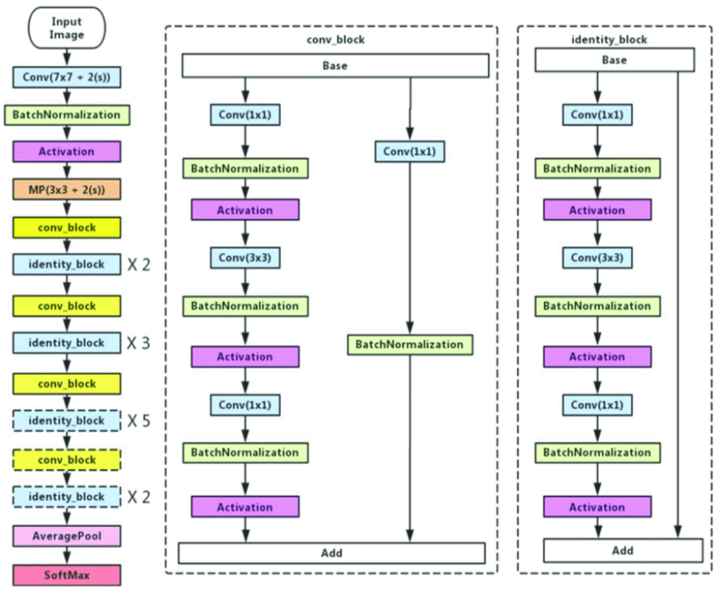

Figure 4. Model ResNet50

Referring to Figure 4, the classification process is performed by ResNet50 with the activation of the fully connected Softmax layer. However, to do classification with SVM, fully connected will be removed and bound with SVM machine learning. So that the model to be built will be like in Figure 5.

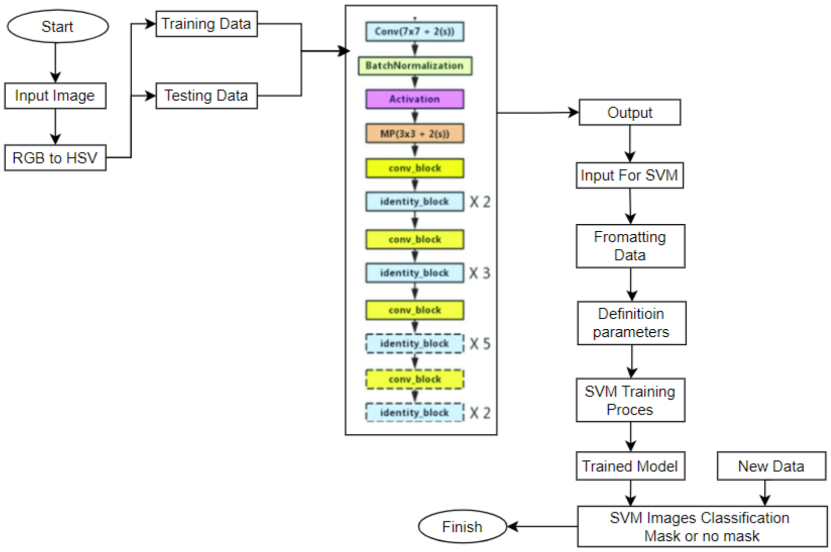

Figure 5. Model Hybrid RGB to HSV – ResNet50 – SVM

In Figure 5, the hybrid model is described starting from the input facial image data, followed by preprocessing, which involves converting the RGB image to HSV. Next, the HSV image dataset is split into two parts, with 75% for training data and 25% for testing data. After the feature extraction process is performed using the ResNet50 model, it is passed on for training using the SVM method.

In the evaluation stage, the experimental results will be summarized, and analysis will be conducted through the creation of graphs and tables. This analysis is used to demonstrate the success of the experiments conducted by the ResNet50-SVM hybrid model, making it applicable for further research.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precission = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1\ Score = \frac{2\ .\ Precission\ .\ Recall}{Precission+Recall} \quad (8)$$

Where TP represents the number of True Positive samples, TN represents the number of True Negative samples, FP represents the number of False Positive samples, and FN represents the number of False Negative samples in the confusion matrix.

3. RESULTS AND DISCUSSION

In the training and testing process, there are a total of 7328 facial images with or without masks. This dataset is divided into 75% training data, which is 5496 images, and 25% testing data, which is 1832 images. The testing strategy involves four approaches: 1) The first approach involves preprocessing the input images by converting them from RGB to HSV and using the ResNet50 model. 2) The second approach involves preprocessing the input images by converting them from RGB to HSV and using the ResNet50-SVM model. Here are some sample data with and without masks.

Figure 6. Sample data without mask

Figure 7. Sample data with mask

3.1. Performance Model RGB – ResNet50

During the training process of 50 epochs, the accuracy of the training data reached 92.36%. Meanwhile, the validation accuracy, reached 86.52.18%. The movement of the accuracy (blue line) and validation accuracy (yellow line) from epoch 1 to 50 can be seen in Figure 8.

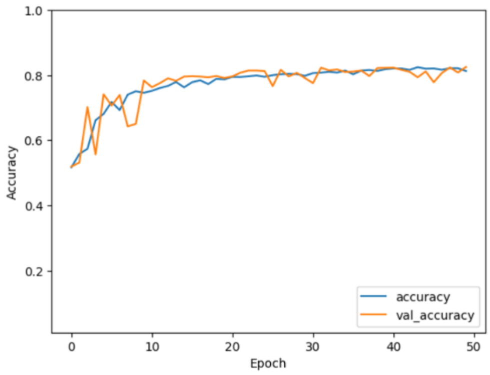

Figure 8. Accuracy and Validation Accuracy Model RGB – ResNet50

3.2. Performance Model RGB to HSV – ResNet50

During the training process of 50 epochs, the accuracy of the training data reached 99.98%. Meanwhile, the validation accuracy, or accuracy on the test data, reached 99.18%. The movement of the accuracy (blue line) and validation accuracy (yellow line) from epoch 1 to 50 can be seen in Figure 9.

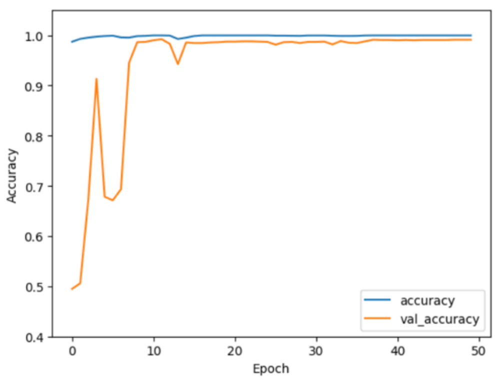

Figure 9 Accuracy and Validation Accuracy Model RGB to HSV – ResNet50

3.3. Performance Model Image RGB – ResNet50 – SVM

In this test, RGB images were used without preprocessing in the facial image dataset. The test results were obtained using a confusion matrix, where the trained model for the test data had TP=829, FP=84, FN=82, and TN=842, as shown in the figure 10.

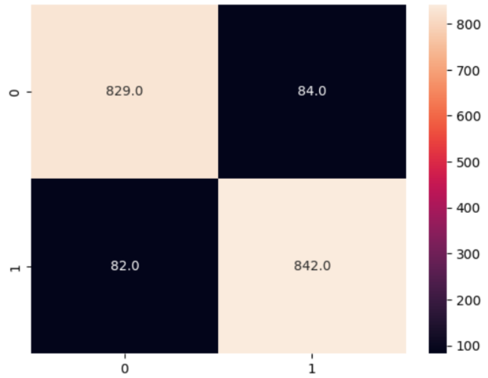

Figure 10. Confusion Matrix RGB – ResNet50 & SVM

From the confusion matrix above, Accuracy, Precision, Recall, and F1 Score can be determined as in the calculation below

a. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{829 + 842}{829 + 842 + 84 + 82} = 0.9086 = 90.86\%$$

b. Precission

$$Precission = \frac{TP}{TP + FP} = \frac{829}{829 + 84} = 0.9079 = 90.79\%$$

c. *Recall*

$$Recall = \frac{TP}{TP + FN} = \frac{829}{829 + 82} = 0.9099 = 90.99\%$$

d. F1 *Score*

$$F1\ Score = \frac{2 * Precission * Recall}{Precission + Recall} = \frac{2 * 0.9079 * 0.9099}{0.9079 + 0.9099} = 0.9088 = 90.88\%$$

3.4. Performance Model RGB to HSV – ResNet50 – SVM

In this test, RGB images were preprocessed into HSV images in the facial image dataset. The test results were obtained using a confusion matrix, where the trained model for the test data had TP=890, FP=36, FN=34, and TN=877, as shown in Figure 11.

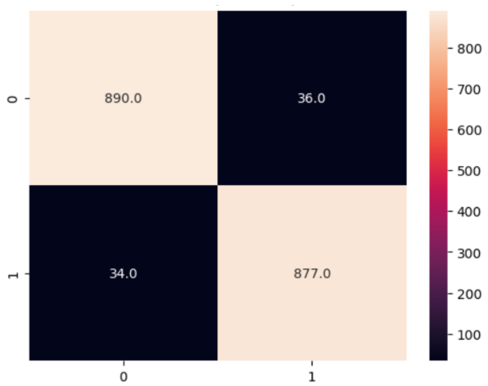

Figure 11. Confusion Matrix RGB to HSV – ResNet50 & SVM

From the confusion matrix above, Accuracy, Precision, Recall, and F1 Score can be determined as in the calculation below

a. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{890 + 877}{890 + 877 + 36 + 34} = 0.9618 = 98.36\%$$

b. Precission

$$Precission = \frac{TP}{TP + FP} = \frac{890}{890 + 36} = 0.9611 = 98.11\%$$

c. Recall

$$Recall = \frac{TP}{TP + FN} = \frac{890}{890 + 34} = 0.9632 = 98.32\%$$

d. F1 Score

$$F1\,Score = \frac{2 * Precission * Recall}{Precission + Recall} = \frac{2 * 0.9611 * 0.9632}{0.9611 + 0.9632} = 0.9028 = 98.21\%$$

From the conducted test, it can be observed that the conversion from RGB to HSV during the preprocessing stage significantly improves the accuracy of the test data. This comparison can be seen in Table 1.

Table 1. Comparation Performance ResNet50 with ResNet50 – SVM

Data Image	Model/Metode	Accuracy
RGB	ResNet50	95.37%
HSV	**ResNet50**	**99.18%**
RGB	ResNet50 – SVM	90.86%
HSV	ResNet50 – SVM	98.36%

In Table 1, the highest accuracy of 99.18% was achieved by the ResNet50 model when the input data underwent preprocessing by converting RGB to HSV. The Hybrid model of transfer learning ResNet50 with SVM achieved an accuracy of 96.18%. For both models, the best results were obtained when the preprocessing step of converting RGB to HSV was performed.

The methods employed in this experiment, namely RGB to HSV - ResNet50 and RGB to HSV - ResNet50 - SVM, will be compared with previous research. The comparison results are presented in Table 2.

Table 2. Comparation ResNet50 & Hybrid ResNet50 – SVM with Related Research Studies

Research	Data Similarity	Model/Metode	Accuracy
[8]	Not Same	ResNet50 – SVM	99.49%
*	**Same +**	**RGB to HSV – ResNet50**	**99.18%**
[7]	Same	MobileNet – GPB	99%
*	**Same +**	**RGB to HSV – ResNet50 – SVM**	**98.38%**
[11]	Not Same	MobileNet	97.75%
[5]	Same	MobileNet – SVM	97.11%
[10]	Not Same	CNN	96.90%
[5]	Same	VGG19 – KNN	96.65%
[5]	Same	MobileNetV2 – KNN	94.92%
[6]	Same	Exception – SVM	94.57%
[9]	Not Same	MobileNetV2 – SSD	91.70%
	Not Same	MobileNetV2	82.96%

In Table 2, RGB to HSV - ResNet50 ranks second in terms of accuracy. However, when comparing it with previous research that used the same data and included additional data, RGB to HSV - ResNet50 achieves the highest accuracy with 99.18%, surpassing MobileNet - GPB with an accuracy of 99%. Meanwhile, the RGB to HSV - Hybrid ResNet50 & SVM model ranks 4th. And in the third position, when using the same data as the researcher, it achieves an accuracy of 98.36%.

From the results of this experiment, we can conclude that the addition of training data affects the accuracy of results. Preprocessing the image data from RGB to HSV before training has a significant impact on improving accuracy. Additionally, performing unfreezing on the ResNet50 model also has an influence on both the RGB to HSV - ResNet50 model and the Hybrid ResNet50 - SVM model.

Next, implementation is carried out using Flask Python to enable real-time prediction of whether a person is wearing a mask or not, based on video frames through a web-based interface. The implementation results are shown in Figure 12, where the person is not wearing a mask.

Figure 12. Video frame without mask detected

As for the implementation of the results using a mask, as seen in Figure 13.

Figure 13. Video frame with mask detected

4. CONCLUSION

From the test results and method discussion for mask detection on facial images, the following conclusions can be drawn: The best model with an accuracy rate of 99.18% is the Transfer Learning ResNet50 model that undergoes preprocessing on the input image by converting RGB to HSV. Preprocessing the image data by converting the colorspace from RGB to HSV in both the ResNet50 transfer learning model and the hybrid ResNet50 with SVM model can improve the accuracy of the trained models. Therefore, the highest accuracy models can be implemented for mask detection.

ACKNOWLEDGEMENTS

The author would like to express gratitude to the professors of Maulana Malik Ibrahim State Islamic University who have taught and guided throughout the research, enabling its completion.

REFERENCES

[1] P. Boldog, T. Tekeli, Z. Vizi, A. Dénes, F. A. Bartha, and G. Röst, “Risk assessment of novel coronavirus COVID-19 outbreaks outside China,” *J Clin Med*, vol. 9, no. 2, Feb. 2020, doi: 10.3390/jcm9020571.

[2] Nurkholis, “Dampak Pandemi Novel-Corona Virus Disiase (Covid-19) Terhadap Psikologi Dan Pendidikan Serta Kebijakan Pemerintah,” *Jurnal PGSD*, vol. Volume 6 (1) 2020, 2020.

[3] O. Russakovsky, J. Deng, Z. Huang, A. C. Berg, and L. Fei-Fei, “Detecting avocados to zucchinis: what have we done, and where are we going?,” 2015. [Online]. Available: www.image-net.org/challenges/LSVRC/2012/analysis/

[4] C. Cortes, V. Vapnik, and L. Saitta, “Support-Vector Networks Editor,” Kluwer Academic Publishers, 1995.

[5] A. Oumina, N. El Makhfi, and M. Hamdi, “Control the COVID-19 Pandemic: Face Mask Detection Using Transfer Learning,” in *2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science, ICECOCS 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020. doi: 10.1109/ICECOCS50124.2020.9314511.

[6] S. Yadav, “Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID-19 Safety Guidelines Adherence,” *Int J Res Appl Sci Eng Technol*, vol. 8, no. 7, pp. 1368–1375, Jul. 2020, doi: 10.22214/ijraset.2020.30560.

[7] I. B. Venkateswarlu, J. Kakarla, and S. Prakash, “Face mask detection using MobileNet and global pooling block,” in *4th IEEE Conference on Information and Communication Technology, CICT 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020. doi: 10.1109/CICT51604.2020.9312083.

[8] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, “A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic,” *Measurement (Lond)*, vol. 167, Jan. 2021, doi: 10.1016/j.measurement.2020.108288.

[9] S. A. Sanjaya and S. A. Rakhmawan, “Face Mask Detection Using MobileNetV2 in the Era of COVID-19 Pandemic,” in *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy, ICDABI 2020*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020. doi: 10.1109/ICDABI51230.2020.9325631.

[10] A. Negi, P. Chauhan, K. Kumar, and R. S. Rajput, “Face Mask Detection Classifier and Model Pruning with Keras-Surgeon,” in *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2020 - Proceeding*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020. doi: 10.1109/ICRAIE51050.2020.9358337.

[11] K. Suresh, M. B. Palangappa, and S. Bhuvan, “Face Mask Detection by using Optimistic Convolutional Neural Network,” in *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, Institute of Electrical and Electronics Engineers Inc., Jan. 2021, pp. 1084–1089. doi: 10.1109/ICICT50816.2021.9358653.

[12] S. Sethi, M. Kathuria, and T. Kaushik, “Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread,” *J Biomed Inform*, vol. 120, Aug. 2021, doi: 10.1016/j.jbi.2021.103848.

[13] G. Moreira, S. A. Magalhães, T. Pinho, F. N. Dos Santos, and M. Cunha, “Benchmark of Deep Learning and a Proposed HSV Colour Space Models for the Detection and Classification of Greenhouse Tomato,” *Agronomy*, vol. 12, no. 2, Feb. 2022, doi: 10.3390/agronomy12020356.

[14] T. H. Nguyen, T. N. Nguyen, and B. V. Ngo, “A VGG-19 Model with Transfer Learning and Image Segmentation for Classification of Tomato Leaf Disease,” *AgriEngineering*, vol. 4, no. 4, pp. 871–887, Dec. 2022, doi: 10.3390/agriengineering4040056.

[15] Prajna Bhandary, “Mask Classifier,” *https://github.com/prajnasb/observations*, 2020.

[16] Vijay Kumar, “Face Mask Detection,” *https://www.kaggle.com/datasets/vijaykumar1799/face-mask-detection*, 2021.